

ℓ_p -norm based James-Stein estimation with minimaxity and sparsity

Yuzo Maruyama

University of Tokyo

e-mail: maruyama@sis.u-tokyo.ac.jp

Abstract: A new class of minimax Stein-type shrinkage estimators of a multivariate normal mean is studied where the shrinkage factor is based on an ℓ_p norm. The proposed estimators allow some but not all coordinates to be estimated by 0 thereby allow sparsity as well as minimaxity.

AMS 2000 subject classifications: Primary 62C20; secondary 62J07.

Keywords and phrases: James-Stein estimator, minimaxity, sparsity.

1. Introduction

Let $Z \sim N_d(\theta, \sigma^2 I_d)$. We are interested in estimation of the mean vector θ with respect to the quadratic loss function $L(\delta, \theta) = \sum_{i=1}^d (\delta_i - \theta_i)^2 / \sigma^2$. Obviously the risk of z is d . We shall say one estimator is as good as the other if the former has a risk no greater than the latter for every θ . Moreover, one dominates the other if it is as good as the other and has smaller risk for some θ . In this case, the latter is called inadmissible. Note that z is a minimax estimator, that is, it minimizes $\sup_{\theta} E[L(\delta, \theta)]$ among all estimators δ . Consequently any δ is as good as z if and only if it is minimax.

Stein (1956) showed that z is inadmissible when $d \geq 3$. James and Stein (1961) explicitly found a class of minimax estimators

$$\hat{\theta}_{\text{JS}} = \left(1 - \frac{c\sigma^2}{\|z\|_2^2}\right) z$$

with $0 \leq c \leq 2(d-2)$ and $\|z\|_2^2 = \sum_{i=1}^d z_i^2$. Baranchik (1964) proposed the James-Stein positive-part estimator

$$\hat{\theta}_{\text{JS}}^+ = \max\left(0, 1 - \frac{c\sigma^2}{\|z\|_2^2}\right) z \quad (1.1)$$

with $0 < c \leq 2(d-2)$ which dominates the James-Stein estimator. A problem with the James-Stein positive-part estimator is, however, that it selects only between two models: the origin and the full model. Zhou and Hwang (2005) overcome the difficulty by utilizing the so-called ℓ_p -norm given by

$$\|z\|_p = \left\{ \sum_{i=1}^d |z_i|^p \right\}^{1/p} \quad (1.2)$$

and in fact proposed minimax estimators $\hat{\theta}_{\text{ZH}}^+$ with the i -th component given by

$$\hat{\theta}_{\text{ZH}}^+ = \max \left(0, 1 - \frac{c\sigma^2}{\|z\|_{2-\alpha}^{2-\alpha} |z_i|^\alpha} \right) z_i \quad (1.3)$$

where $0 \leq \alpha < (d-2)/(d-1)$ and $0 < c \leq 2\{(d-2) - \alpha(d-1)\}$. When $\alpha > 0$, the i -th component of the estimator with

$$\frac{|z_i|}{\sigma} \leq c^{1/\alpha} \left(\frac{\sigma}{\|z\|_{2-\alpha}} \right)^{(2-\alpha)/\alpha} \quad (1.4)$$

becomes zero. Hence the choice between a full model and reduced models, where some coefficients are reduced to zero, is possible.

In this paper, we establish minimaxity of a new class of ℓ_p -norm based shrinkage estimators $\hat{\theta}_{\text{LP}}^+$ with the i -th component given by

$$\hat{\theta}_{\text{LP}}^+ = \max \left(0, 1 - \frac{c\sigma^2}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i \quad (1.5)$$

where $0 \leq \alpha < (d-2)/(d-1)$, $p > 0$, $0 < c \leq 2(d-2)\gamma(d, p, \alpha)$ and

$$\gamma(d, p, \alpha) = \min(1, d^{(2-p-\alpha)/p}) \left\{ 1 - \alpha \frac{d-1}{d-2} \right\}.$$

When α is strictly positive in (1.5), sparsity happens as in (1.4). In [Zhou and Hwang \(2005\)](#), $p = 2 - \alpha$ was assumed and the ℓ_p -norm with

$$d/(d-1) < p = 2 - \alpha < 2$$

was treated. From their proof, the choice of $p = 2 - \alpha$ seemed only applicable for constructing estimators with minimaxity and sparsity simultaneously. We produce such minimax estimators based on the ℓ_p -norm for all $p > 0$. As an extreme case ($p = \infty$), we can show that

$$\max \left(0, 1 - \sigma^2 \frac{2\{(d-2) - \alpha(d-1)\}}{d \{\max |z_i|\}^{2-\alpha} |z_i|^\alpha} \right) z_i$$

with $0 \leq \alpha < (d-2)/(d-1)$ is minimax. A more general result of minimaxity, corresponding to the result of [Efron and Morris \(1976\)](#), where c is replaced by $\phi(\|z\|_p/\sigma)$ in (1.5), is given in Section 2. In Section 3, the corresponding results for unknown σ^2 are presented.

2. Minimaxity with sparsity: known scale

In this section, we assume that σ^2 is known and establish conditions under which estimators $\hat{\theta}_\phi$ of the form

$$\hat{\theta}_{i\phi} = \left(1 - \frac{\sigma^2 \phi(\|z\|_p/\sigma)}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i \quad (2.1)$$

as the i -th component, are minimax. Note the shrinkage factor of (2.1), $1 - \sigma^2 \phi(\|z\|_p/\sigma) / \{\|z\|_p^{2-\alpha} |z_i|^\alpha\}$ is symmetric with respect to z_i . As shown in Theorem 4 of Zhou and Hwang (2005), the shrinkage estimator with the symmetry is dominated by the positive-part estimator. Hence the minimaxity of $\hat{\theta}_\phi^+$ follows from the minimaxity of $\hat{\theta}_\phi$.

Recall that the risk of z is equal to d or finite. Hence a straightforward application of Schwarz's inequality shows that the risk of $z + \xi(z)$ is finite if and only if

$$E \left[\sum_{i=1}^d \{\xi_i(z)\}^2 / \sigma^2 \right] < \infty. \quad (2.2)$$

In that case, Stein's (1981) identity states that if $\xi(z)$ is absolutely continuous, we have

$$E[(z_i - \theta_i)\xi(z)] = \sigma^2 E[(\partial/\partial z_i)\xi(z)] \quad (2.3)$$

for $i = 1, \dots, d$ and each expectation exists.

In this paper, we assume $0 \leq \alpha < 1$ and ϕ is bounded, say $|\phi| \leq M$ for some $M > 0$. Under these assumptions, (2.2) follows with $\xi = (\xi_1, \dots, \xi_d)'$ and

$$\xi_i(z) = \hat{\theta}_{i\phi} - z_i = -\frac{\sigma^2 \phi(\|z\|_p/\sigma)}{\|z\|_p^{2-\alpha} |z_i|^\alpha} z_i. \quad (2.4)$$

In fact, we have

$$\frac{\sum_{i=1}^d \{\xi_i(z)\}^2}{\sigma^2} = \sigma^2 \phi(\|z\|_p/\sigma)^2 \frac{\|z\|_{2(1-\alpha)}^{2(1-\alpha)}}{\|z\|_p^{2(2-\alpha)}} \leq \sigma^2 M^2 \frac{\|z\|_{2(1-\alpha)}^{2(1-\alpha)}}{\|z\|_p^{2(2-\alpha)}}$$

and further

$$\frac{\|z\|_{2(1-\alpha)}^{2(1-\alpha)}}{\|z\|_p^{2(2-\alpha)}} \leq \frac{\max(1, d^{(p-2+2\alpha)/\{2p(1-\alpha)\}})}{\|z\|_{2(1-\alpha)}^2} \leq \frac{\max(1, d^{(p-2+2\alpha)/\{2p(1-\alpha)\}})}{\|z\|_2^2}$$

by Part 1 of Lemma A.1 in Appendix. Since $E[\sigma^2/\|z\|_2^2] \leq 1/(d-2)$, for ξ given by (2.4) we have

$$E \left[\sum_{i=1}^d \{\xi_i(z)\}^2 / \sigma^2 \right] \leq \frac{M^2 \max(1, d^{(p-2+2\alpha)/\{2p(1-\alpha)\}})}{d-2}.$$

Hence under the assumption of bounded ϕ , the risk of $\hat{\theta}_\phi$ given by (2.1) is finite. Further, with an additional assumption that ϕ is absolutely continuous, Stein's (1981) identity given by (2.3) is available for derivation of Stein's (1981) unbiased risk estimator.

Lemma 2.1. *Assume that $\phi(v)$ is bounded and absolutely continuous and that $0 \leq \alpha < 1$.*

1. The risk function of the estimator $\hat{\theta}_\phi$ is

$$E \left[\frac{\|\hat{\theta}_\phi - \theta\|_2^2}{\sigma^2} \right] = d + E \left[\sum_i \left(\frac{|z_i|}{\|z\|_p} \right)^{p-\alpha} \frac{\phi(\|z\|_p/\sigma) \psi_\phi(z/\sigma)}{(\|z\|_p/\sigma)^2} \right] \quad (2.5)$$

where

$$\begin{aligned} \psi_\phi(z) &= \frac{\phi(\|z\|_p)}{\|z\|_p^{-p-\alpha+2}} \frac{\sum_i |z_i|^{2(1-\alpha)}}{\sum_i |z_i|^{p-\alpha}} - 2(1-\alpha) \|z\|_p^p \frac{\sum_i |z_i|^{-\alpha}}{\sum_i |z_i|^{p-\alpha}} \\ &\quad - 2 \left\{ \alpha - 2 + \|z\|_p \frac{\phi'(\|z\|_p)}{\phi(\|z\|_p)} \right\}. \end{aligned} \quad (2.6)$$

2. Assume $\phi(v) \geq 0$. Then $\psi_\phi(z) \leq \Psi_\phi(\|z\|_p)$ where

$$\Psi_\phi(v) = \max(1, d^{(p+\alpha-2)/p}) \phi(v) - 2 \{d - 2 - \alpha(d-1)\} - 2 \frac{v\phi'(v)}{\phi(v)}. \quad (2.7)$$

Proof. From the invariance with respect to the transformation, $z \rightarrow cz$, we can take $c = 1/\sigma$ and hence, without the loss of generality, assume $\sigma^2 = 1$ in the proof.

[Part 1] Let $v = \|z\|_p$. Componentwisely we have

$$\begin{aligned} (\hat{\theta}_i - \theta_i)^2 &= \{(1 - \phi(v)v^{\alpha-2}|z_i|^{-\alpha}) z_i - \theta_i\}^2 \\ &= (z_i - \theta_i)^2 + \phi^2(v)v^{2(\alpha-2)}|z_i|^{2(1-\alpha)} - 2(z_i - \theta_i) \{\phi(v)v^{\alpha-2}|z_i|^{-\alpha} z_i\}. \end{aligned} \quad (2.8)$$

For the third term of the right-hand side of (2.8), the Stein identity given by (2.3) is applicable. Note

$$\frac{\partial}{\partial z_i} v = v^{1-p}|z_i|^{p-2} z_i, \quad \frac{\partial}{\partial z_i} \{|z_i|^{-\alpha} z_i\} = (1-\alpha)|z_i|^{-\alpha}. \quad (2.9)$$

Then the differentiation of $\phi(v)v^{\alpha-2}|z_i|^{-\alpha} z_i$ with respect to z_i is given by

$$\begin{aligned} &(1-\alpha)\phi(v)v^{\alpha-2}|z_i|^{-\alpha} + (\alpha-2)\phi(v)v^{\alpha-p-2}|z_i|^{p-\alpha} + \phi'(v)v^{\alpha-p-1}|z_i|^{p-\alpha} \\ &= \{\phi(v)v^{\alpha-p-2}\} \{(1-\alpha)v^p|z_i|^{-\alpha} + \{(\alpha-2) + v\phi'(v)/\phi(v)\}|z_i|^{p-\alpha}\} \end{aligned}$$

and Part 1 follows by taking summation with respect to i .

[Part 2] Recall $0 \leq \alpha < 1$ and $p > 0$. By Part 2 of Lemma A.1 in Appendix, we have

$$\sum_{i=1}^d |z_i|^{-\alpha} \geq d \frac{\sum_{i=1}^d |z_i|^{p-\alpha}}{\sum_{i=1}^d |z_i|^p} = d \frac{\sum_{i=1}^d |z_i|^{p-\alpha}}{\|z\|_p^p} \quad (2.10)$$

and, by Part 3 of Lemma A.1,

$$\frac{1}{\|z\|_p^{-p-\alpha+2}} \frac{\sum_i |z_i|^{2(1-\alpha)}}{\sum_i |z_i|^{p-\alpha}} = \frac{\sum_i s_i^{2(1-\alpha)/p}}{\sum_i s_i^{(p-\alpha)/p}} \leq \max(1, d^{(p+\alpha-2)/p}) \quad (2.11)$$

where $s_i = |z_i|^p / \|z\|_p^p$ with $\sum_{i=1}^d s_i = 1$ and $s_i \geq 0$ for any i . By applying these inequalities to (2.6), Part 2 follows. \square

By Lemma 2.1, a sufficient condition for $E[\|\hat{\theta} - \theta\|_2^2] \leq d$ is

$$\Psi_\phi(v) \leq 0 \quad (2.12)$$

as well as the assumption of Lemma 2.1. When ϕ is monotone non-decreasing, we easily have a following result for minimaxity, which corresponds to the result by Baranchik (1970) with $\alpha = 0$ and $p = 2$.

Theorem 2.1. *Assume $d \geq 3$ and $0 \leq \alpha < (d-2)/(d-1)$. Assume $\phi(v)$ is absolutely continuous, monotone non-decreasing and*

$$0 \leq \phi(v) \leq 2(d-2)\gamma(d, p, \alpha)$$

where $\gamma(d, p, \alpha)$ is given by

$$\gamma(d, p, \alpha) = \min(1, d^{(2-p-\alpha)/p}) \left\{ 1 - \alpha \frac{d-1}{d-2} \right\}. \quad (2.13)$$

Under known σ^2 , the shrinkage estimator $\hat{\theta}_\phi$, with the i -th component,

$$\hat{\theta}_{i\phi} = \left(1 - \frac{\sigma^2 \phi(\|z\|_p/\sigma)}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i$$

is minimax.

More generally, by the derivative,

$$\begin{aligned} & \frac{d}{dv} \left\{ \frac{v^b \phi(v)}{\{a - \phi(v)\}^c} \right\} \\ &= \frac{bv^{b-1} \phi(v)}{\{a - \phi(v)\}^{c+1}} \left(\frac{c-1}{b} v \phi'(v) + \frac{a}{b} \frac{v \phi'(v)}{\phi(v)} + a - \phi(v) \right), \end{aligned} \quad (2.14)$$

we have a following sufficient condition as in Efron and Morris (1976).

Theorem 2.2. *Assume $d \geq 3$ and $0 \leq \alpha < (d-2)/(d-1)$. Assume $\phi(v)$ is absolutely continuous and*

$$0 \leq \phi(v) \leq 2(d-2)\gamma(d, p, \alpha).$$

Further, for all v with $\phi(v) < 2(d-2)\gamma(d, p, \alpha)$

$$g_\phi(v) = \frac{v^{d-2-\alpha(d-1)} \phi(v)}{2(d-2)\gamma(d, p, \alpha) - \phi(v)}$$

is assumed to be non-decreasing. Further if there exists $v_* > 0$ such that $\phi(v) = 2(d-2)\gamma(d, p, \alpha)$, then $\phi(v)$ is assumed equal to $2(d-2)\gamma(d, p, \alpha)$ for all $v \geq v_*$. Then $\hat{\theta}_\phi$ is minimax.

Recall that ℓ_p norm with any positive p is available in Lemma 2.1 and Theorem 2.2. As an extreme case ($p = \infty$), we have $\lim_{p \rightarrow \infty} \gamma(d, p, \alpha) = \{1 - \alpha(d - 1)/(d - 2)\}/d$ and hence

$$\max \left(0, 1 - \sigma^2 \frac{2\{(d-2) - \alpha(d-1)\}}{d \{\max |z_i|\}^{2-\alpha} |z_i|^\alpha} \right) z_i$$

with $0 \leq \alpha < (d-2)/(d-1)$ is minimax.

Remark 2.1. The solution of $\Psi_\phi(v) = 0$ or $g_\phi(v) = 1/\lambda$ for any $\lambda > 0$, is

$$\phi_{\text{DS}}(v) = \frac{2(d-2)\gamma(d, p, \alpha)}{1 + \lambda v^{d-2-\alpha(d-1)}},$$

under which Dasgupta and Strawderman (1997) showed the risk of the estimator with $\phi_{\text{DS}}(v)$ is exactly equal to d when $p = 2$ and $\alpha = 0$. Actually it is related to the concept of “near unbiasedness” or “approximate unbiasedness” in the literature of SCAD (smoothly clipped absolute deviation) including Antoniadis and Fan (2001). Since $\phi_{\text{DS}}(v)$ is monotone decreasing and approaches 0 as $v \rightarrow \infty$, unnecessary modeling biases are effectively avoided with $\phi_{\text{DS}}(v)$.

3. Minimavity with sparsity: unknown scale

In this section, we assume that σ^2 is unknown and that $S \sim \sigma^2 \chi_n^2$ is additionally observed. We establish minimavity result of the shrinkage estimators $\hat{\theta}_\phi$ with the i -th component given by

$$\begin{aligned} \hat{\theta}_{i\phi} &= \left(1 - \frac{\hat{\sigma}^2 \phi(\|z\|_p / \sqrt{\hat{\sigma}^2})}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i \\ &= \left(1 - \frac{s}{n+2} \frac{\phi(\sqrt{n+2} \|z\|_p / \sqrt{s})}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i \end{aligned} \quad (3.1)$$

where $\hat{\sigma}^2 = s/(n+2)$.

Lemma 3.1. *Assume that $\phi(u)$ is, non-negative, bounded and absolutely continuous and that $0 \leq \alpha < 1$. Then the risk function of the estimator $\hat{\theta}_\phi$ is*

$$E \left[\frac{\|\hat{\theta}_\phi - \theta\|_2^2}{\sigma^2} \right] \leq d + E \left[\sum_i \left\{ \frac{|z_i|}{\|z\|_p} \right\}^{p-\alpha} \frac{\phi(u)}{u^2} \left(\Psi_\phi(u) - \frac{2u\phi'(u)}{n+2} \right) \right] \quad (3.2)$$

where $u = \|z\|_p / \sqrt{\hat{\sigma}^2}$ and $\Psi_\phi(u)$ is given by (2.7).

Proof. From the invariance with respect to the transformation, $z \rightarrow cz$ and $s \rightarrow c^2 s$, we can take $c = 1/\sigma$ and hence, without the loss of generality, $\sigma^2 = 1$

is assumed in the proof. Let $v = \|z\|_p$ and $u = v/\sqrt{\hat{\sigma}^2}$. Componentwisely we have

$$\begin{aligned} (\hat{\theta}_i - \theta_i)^2 &= \left\{ \left(1 - \frac{\phi(u)\hat{\sigma}^2}{v^{2-\alpha}|z_i|^\alpha} \right) z_i - \theta_i \right\}^2 \\ &= (z_i - \theta_i)^2 + \frac{\phi^2(u)\{\hat{\sigma}^2\}^2}{v^{2(2-\alpha)}}|z_i|^{2(1-\alpha)} - 2\hat{\sigma}^2(z_i - \theta_i)\frac{\phi(u)z_i}{v^{2-\alpha}|z_i|^\alpha} \end{aligned} \quad (3.3)$$

and hence

$$\begin{aligned} \sum_{i=1}^d (\hat{\theta}_i - \theta_i)^2 &= \sum_{i=1}^d (z_i - \theta_i)^2 + \frac{\phi^2(u)\{\hat{\sigma}^2\}^2}{v^{2(2-\alpha)}} \sum_{i=1}^d |z_i|^{2(1-\alpha)} \\ &\quad - 2\hat{\sigma}^2 \sum_{i=1}^d (z_i - \theta_i) \left\{ \phi(u)v^{\alpha-2}|z_i|^{-\alpha} z_i \right\}. \end{aligned} \quad (3.4)$$

For the third term of the right-hand side of (3.4), the Stein identity given by (2.3) is applicable. By (2.9), the differentiation of $\phi(v/\sqrt{\hat{\sigma}^2})v^{\alpha-2}|z_i|^{-\alpha}z_i$ with respect to z_i is

$$\begin{aligned} &\frac{(1-\alpha)\phi(v/\sqrt{\hat{\sigma}^2})}{v^{2-\alpha}}|z_i|^{-\alpha} + \frac{(\alpha-2)\phi(v/\sqrt{\hat{\sigma}^2})}{v^{p+2-\alpha}}|z_i|^{p-\alpha} + \frac{\phi'(v/\sqrt{\hat{\sigma}^2})}{\sqrt{\hat{\sigma}^2}v^{p+1-\alpha}}|z_i|^{p-\alpha} \\ &= \frac{\phi(u)}{v^{p+2-\alpha}} \left((1-\alpha)v^p|z_i|^{-\alpha} + \left\{ (\alpha-2) + u\frac{\phi'(u)}{\phi(u)} \right\} |z_i|^{p-\alpha} \right). \end{aligned}$$

By the inequality (2.10) and the Stein identity, we have

$$\begin{aligned} &-2E \left[\hat{\sigma}^2 \sum_{i=1}^d (z_i - \theta_i) \left\{ \phi(v/\sqrt{\hat{\sigma}^2})v^{\alpha-2}|z_i|^{-\alpha} z_i \right\} \right] \\ &\leq -E \left[\frac{\hat{\sigma}^2 \sum_i |z_i|^{p-\alpha}}{v^{2-\alpha+p}} \phi(u) \left(2(d-2) - \alpha(d-1) + 2u\frac{\phi'(u)}{\phi(u)} \right) \right] \\ &= -E \left[\sum_i \left\{ \frac{|z_i|}{\|z\|_p} \right\}^{p-\alpha} \frac{\phi(u)}{u^2} \left(2(d-2) - \alpha(d-1) + 2u\frac{\phi'(u)}{\phi(u)} \right) \right]. \end{aligned}$$

For the second term of the right-hand side of (3.4), a well known identity for chi-square distributions (see e.g. [Efron and Morris \(1976\)](#))

$$E[sh(s)] = \sigma^2 E[nh(s) + 2sh'(s)] \quad (3.5)$$

for $s \sim \sigma^2 \chi_n^2$ is applicable. The differentiation of

$$\frac{\phi^2(v/\sqrt{\hat{\sigma}^2})\{\hat{\sigma}^2\}^2}{s} = \frac{\phi^2(\sqrt{n+2}v/\sqrt{s})s}{(n+2)^2},$$

with respect to s , is

$$\begin{aligned} & \frac{\phi^2(\sqrt{n+2}v/\sqrt{s}) - \phi(\sqrt{n+2}v/\sqrt{s})\phi'(\sqrt{n+2}v/\sqrt{s})\sqrt{n+2}v/s^{1/2}}{(n+2)^2} \\ &= \frac{\phi^2(u) - u\phi(u)\phi'(u)}{(n+2)^2}. \end{aligned} \quad (3.6)$$

Hence, by the identity (3.5) with (3.6), we have

$$E_{s|v} [\phi^2(u)\{\hat{\sigma}^2\}^2] = E_{s|v} \left[\hat{\sigma}^2 \phi(u) \left\{ \phi(u) - \frac{2}{n+2} u\phi'(u) \right\} \right]. \quad (3.7)$$

Further, by (2.11) and (3.7), we have

$$\begin{aligned} & E \left[\frac{\phi^2(v/\sqrt{\hat{\sigma}^2})\{\hat{\sigma}^2\}^2}{v^{2(2-\alpha)}} \sum_{i=1}^d |z_i|^{2(1-\alpha)} \right] \\ & \leq \max(1, d^{(p+\alpha-2)/p}) E \left[\frac{\sum_i |z_i|^{p-\alpha} \hat{\sigma}^2}{v^{2-\alpha+p}} \phi(u) \left(\phi(u) - \frac{2}{n+2} u\phi'(u) \right) \right] \\ & = \max(1, d^{(p+\alpha-2)/p}) E \left[\sum_i \left\{ \frac{|z_i|}{\|z\|_p} \right\}^{p-\alpha} \frac{\phi(u)}{u^2} \left(\phi(u) - \frac{2}{n+2} u\phi'(u) \right) \right]. \end{aligned}$$

□

By Lemma 3.1, a sufficient condition for $E[\|\hat{\theta} - \theta\|_2^2] \leq d$ is

$$\Psi_\phi(u) - \frac{2u\phi'(u)}{n+2} \leq 0 \quad (3.8)$$

as well as the assumptions of Lemma 3.1. When ϕ is monotone non-decreasing, as in Theorem 2.1 for the known scale case, we easily have a following result for minimaxity.

Theorem 3.1. *Assume $d \geq 3$ and $0 \leq \alpha < (d-2)/(d-1)$. Assume $\phi(u)$ is absolutely continuous, monotone non-decreasing and*

$$0 \leq \phi(u) \leq 2(d-2)\gamma(d, p, \alpha)$$

where $\gamma(d, p, \alpha)$ is given by (2.13). Under unknown σ^2 , the shrinkage estimator $\hat{\theta}_\phi$, with the i -th component,

$$\hat{\theta}_{i\phi} = \left(1 - \frac{\hat{\sigma}^2 \phi(\|z\|_p/\sqrt{\hat{\sigma}^2})}{\|z\|_p^{2-\alpha} |z_i|^\alpha} \right) z_i$$

is minimax.

Hence Theorem 3.1 guarantees that Theorem 2.1 remains true if σ^2 is replaced by the estimator $\hat{\sigma}^2 = s/(n+2)$. By following Efron and Morris (1976) and using the relation (2.14), a more general theorem corresponding to Theorem 2.2 is given as follows.

Theorem 3.2. Assume $d \geq 3$ and $0 \leq \alpha < (d-2)/(d-1)$. Assume $\phi(u)$ is absolutely continuous and

$$0 \leq \phi(u) \leq 2(d-2)\gamma(d, p, \alpha)$$

where $\gamma(d, p, \alpha)$ is given by (2.13). Further, for all u with $\phi(u) < 2(d-2)\gamma(d, p, \alpha)$

$$g_\phi(u) = \frac{u^{d-2-\alpha(d-1)}\phi(u)}{\{2(d-2)\gamma(d, p, \alpha) - \phi(u)\}^{1+2\{d-2-\alpha(d-1)\}/(n+2)}}$$

is assumed to be non-decreasing. Further if there exists $u_* > 0$ such that $\phi(u) = 2(d-2)\gamma(d, p, \alpha)$, then $\phi(u)$ is assumed equal to $2(d-2)\gamma(d, p, \alpha)$ for all $u \geq u_*$. Then $\hat{\theta}_\phi$ is minimax.

We see that Theorem 2.2 for known σ^2 guarantees minimaxity of $\hat{\theta}_\phi$ with ϕ which is not monotone non-decreasing. As I mentioned in Remark 2.1, even a monotone decreasing $\phi_{\text{DS}}(v)$, which is the solution $g_\phi(u) = \lambda$, leads minimaxity. In unknown variance case, however, the solution of $g_\phi(u) = \lambda$ in Theorem 3.2, is not tractable. An alternative to $\phi_{\text{DS}}(v)$ is

$$\tilde{\phi}_{\text{DS}}(u) = \frac{2(d-2)\gamma(d, p, \alpha)}{1 + \lambda u^l},$$

where

$$l = \frac{d-2-\alpha(d-1)}{1+2\{d-2-\alpha(d-1)\}/(n+2)},$$

By straightforward calculation, $g_\phi(u)$ with $\tilde{\phi}_{\text{DS}}(u)$ is increasing.

Appendix A: Some inequalities

Here we summarize some inequalities which are used in the main article.

Lemma A.1. 1. Let $q > r > 0$. Then

$$\|z\|_q^r \leq \|z\|_r^r \leq d^{1-r/q} \|z\|_q^r. \quad (\text{A.1})$$

2. Let $q \geq 0$ and $r \geq 0$. Then

$$d \sum_{i=1}^d |z_i|^{q-r} \leq \sum_{i=1}^d |z_i|^{-r} \sum_{i=1}^d |z_i|^q. \quad (\text{A.2})$$

3. Let $a \geq 0$ and $b \leq 1$. Assume $\sum_{i=1}^d s_i = 1$ and $s_i \geq 0$ for all i . Then

$$\frac{\sum_{i=1}^d s_i^a}{\sum_{i=1}^d s_i^b} \leq \max(1, d^{b-a}). \quad (\text{A.3})$$

Proof. [Part 1] In the first inequality, we have

$$\frac{\|z\|_q^r}{\|z\|_r^r} = \left\{ \frac{\|z\|_q^q}{\|z\|_r^q} \right\}^{r/q} = \left(\sum_{i=1}^d \left\{ \frac{|z_i|^r}{\|z\|_r^r} \right\}^{q/r} \right)^{r/q} \leq \left(\sum_{i=1}^d \frac{|z_i|^r}{\|z\|_r^r} \right)^{r/q} = 1$$

since $|z_i|^r/\|z\|_r^r \leq 1$ and $q/r \geq 1$. In the second inequality, let X be a discrete random variable with the probability mass function $\Pr(X = |z_1|^r) = \Pr(X = |z_2|^r) = \cdots = \Pr(X = |z_d|^r) = 1/d$. Then

$$\|z\|_r^r/d = E[X] \leq \left\{ E[X^{q/r}] \right\}^{r/q} = \left\{ \|z\|_q^q/d \right\}^{r/q} = d^{-r/q} \|z\|_q^r$$

where $q/r > 1$ and the inequality is from Jensen's inequality.

[Part 2] Let X be a discrete random variable with the probability mass function $\Pr(X = |z_1|) = \Pr(X = |z_2|) = \cdots = \Pr(X = |z_d|) = 1/d$. Then we have

$$\frac{\sum_{i=1}^d |z_i|^{q-r}}{d} = E[X^{q-r}], \quad \frac{\sum_{i=1}^d |z_i|^q}{d} = E[X^q], \quad \frac{\sum_{i=1}^d |z_i|^{-r}}{d} = E[X^{-r}].$$

From the correlation inequality $E[X^{q-r}] \leq E[X^q]E[X^{-r}]$, the inequality (A.2) follows.

[Part 3] Let $f(\mathbf{s}, c) = \sum_{i=1}^d s_i^c$ with $\mathbf{s} = (s_1, \dots, s_d)$. For any fixed \mathbf{s} , $f(\mathbf{s}, c)$ is non-increasing in c . For $a \geq b$, we have clearly $f(\mathbf{s}, a)/f(\mathbf{s}, b) \leq 1$ and the equality is attained by $\mathbf{s} = (1, 0, \dots, 0)$. When $a < b$, we have $0 \leq a < b \leq 1$ from the assumption and hence

$$d = f(\mathbf{s}, 0) \geq f(\mathbf{s}, a) \geq f(\mathbf{s}, b) \geq f(\mathbf{s}, 1) = 1$$

and $1 \leq f(\mathbf{s}, a)/f(\mathbf{s}, b) \leq d$ for any \mathbf{s} . By the method of Lagrange multiplier, $\hat{\mathbf{s}} = (1, \dots, 1)/d$ gives the maximum value, $f(\hat{\mathbf{s}}, a)/f(\hat{\mathbf{s}}, b) = d^{b-a}$. \square

References

- ANTONIADIS, A. and FAN, J. (2001). Regularization of wavelet approximations. *J. Amer. Statist. Assoc.* **96** 939–967. With discussion and a rejoinder by the authors. [MR1946364](#)
- BARANCHIK, A. J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution Technical Report No. 51, Department of Statistics, Stanford University.
- BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645. [MR0253461](#)
- DASGUPTA, A. and STRAWDERMAN, W. E. (1997). All estimates with a given risk, Riccati differential equations and a new proof of a theorem of Brown. *Ann. Statist.* **25** 1208–1221. [MR1447748](#)
- EFRON, B. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21. [MR0403001](#)

- JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, Calif. [MR0133191](#)
- STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I* 197–206. University of California Press, Berkeley and Los Angeles. [MR0084922](#)
- STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR630098](#)
- ZHOU, H. H. and HWANG, J. T. G. (2005). Minimax estimation with thresholding and its application to wavelet analysis. *Ann. Statist.* **33** 101–125. [MR2157797](#)